

# Cash Flow Entropy

Ulrich Kirchner

PO Box 72277, Parkview, 2122, South Africa  
ulrich.kirchner@gmail.com

Simon Moolman

Department of Physics  
University of the Witwatersrand  
Johannesburg

January 24, 2013

## Abstract

In this paper we aim to find a measure for the diversity of cash flows between agents in an economy. We argue that cash flows can be linked to probabilities of finding a currency unit in a given cash flow. We then use the information entropy as a natural measure of diversity. This leads to a hierarchical inequality measure, which includes the well-known Theil index as a special case, and a constraint on agent entropy differentials, i.e., the difference between cash inflow and outflow entropies. The last result is particularly intriguing as it formalises the fact that an economy must contain both, cash flow aggregating and cash flow diversifying agents.

## 1 Introduction

Most people, when inspecting their bank account statement, could make the following observation: while inflows tend to come in a few large payments (salary etc), outflows are much more dispersed (daily

groceries, rent, monthly bills, etc). By considering the sources, destinations and relative sizes of the payments, it is possible to write down the probability of where a dollar of an agent's income came from and where it is going. With this we can describe the difference between the large inflows and dispersed outflows in terms of the entropy of these probabilities.

Let us assume an economy in which no money is created or destroyed and in which, over a period, cash flows only occur between a set of discrete agents. Such an economy can clearly not exist if every agent's cash flow situation is as above — few large inflows and many small outflows. There must be agents who collect small cash flows and pay out large sums. Examples which come to mind are generally companies, for example selling a mass produced product (small cash flows) and buying supplies and services in bulk (large payments).

This hints at the idea that the entropies of the inflow and outflow probabilities of all the agents in the economy must be related in some way.

This paper will be structured as follows: In Section

2 we formalize the cash flow-probability relationship and introduce information entropy to measure diversity. In Section 3 we examine the results for the simpler case of a steady-state economy. Section 4 contains examples for the two and three agent cases. This is followed by the Conclusion in Section 5. In the appendix we derive the information entropy subdivision property for the case of 2/3 possible outcomes.

## 2 Cash flows and probabilities

Let us assume that in an economy cash flows occur over a period of time between  $N$  agents. Let  $c_k^j$  represent the cash flow from agent  $j$  to agent  $k$ . This can be conveniently represented in a cash flow matrix, where each row corresponds to a source agent (upper index) and each column to a destination agent (lower index)

$$\begin{bmatrix} c_1^1 & \dots & c_N^1 \\ \vdots & & \vdots \\ c_1^N & \dots & c_N^N \end{bmatrix} \quad (1)$$

With cash flow  $c_k^j$  we can then associate a probability

$$p_k^j \stackrel{\text{def}}{=} \frac{c_k^j}{\sum_{k,j} c_k^j}. \quad (2)$$

which represents the likelihood that a random currency unit in the economy would have been part of cash flow  $c_k^j$  (which assumes identifiable currency units). Note that “saving” would correspond to a cashflow  $c_j^j$  from an agent  $j$  to himself.

To measure the diversity of all cash flows in the economy we will use the information entropy

$$H \stackrel{\text{def}}{=} - \sum_{j,k} p_k^j \log_2(p_k^j), \quad (3)$$

which is measured in bits (due to the choice of 2 as the base of the logarithm). To simplify the notation let us extend this definition to “un-normalized” probabilities by defining a new function

$$K(x_1, \dots, x_n) \stackrel{\text{def}}{=} H\left(\frac{x_1}{\sum_i x_i}, \dots, \frac{x_n}{\sum_i x_i}\right). \quad (4)$$

An intriguing property of the information entropy is how it respects the grouping of possible outcomes. Lets say we group the  $N$  possible outcomes into  $n < N$  groups. The overall entropy can then be written as

$$H = H_g + \sum_j p_j H_j, \quad (5)$$

where  $p_j$  is the probability of the outcome being in group  $j$ ,  $H_j$  is the “internal entropy” of group  $j$  (entropy of the probabilities conditional on the fact that the outcome was in group  $j$ ), and

$$H_g \stackrel{\text{def}}{=} - \sum_j p_j \log_2 p_j \quad (6)$$

is the information entropy for the aggregate groups.

By applying above identity recursively to the probabilities corresponding to all possible cash flows one obtains a hirarchical set of inequality/diversity measures.

As a first step let us put all savings  $c_i^i$  in one group, and all remaining cash flows in another group. Using above identity we find

$$H = H_{\{sc\}} + p_{\{s\}} H_{\{s\}} + p_{\{c\}} H_{\{c\}}, \quad (7)$$

where  $p_{\{s\}} \stackrel{\text{def}}{=} \sum_i p_i^i$  and  $p_{\{c\}} = 1 - p_{\{s\}}$  are the aggregate probability of a currency unit being in a “saving cash flow” and inter-agent cash flow respectively,

$$H_{\{sc\}} \stackrel{\text{def}}{=} -p_{\{s\}} \log_2(p_{\{s\}}) - p_{\{c\}} \log_2(p_{\{c\}}) \quad (8)$$

is a measure of diversification between savings and spending,

$$H_{\{s\}} \stackrel{\text{def}}{=} K(c_1^1, \dots, c_N^N) \quad (9)$$

is a measure of the saving diversity in the economy (it is highest when all agents have the same saving), and

$$H_{\{c\}} \stackrel{\text{def}}{=} K(c_2^1, \dots, c_{n-1}^n) \quad (10)$$

is a measure of diversity for all inter-agent cash flows.

Cash flows can now be grouped by their origin or destination. We define the aggregated cash flows *from agent  $j$  and to agent  $k$  excluding savings* as

$$c^j \stackrel{\text{def}}{=} \sum_{k, k \neq j} c_k^j \quad c_k \stackrel{\text{def}}{=} \sum_{j, j \neq k} c_k^j. \quad (11)$$

The probabilities of finding a currency unit (which is not a saving) in the cash-out/in flows of an agent are

$$p^j = \frac{c^j}{\sum_i c^i} \quad p_k = \frac{c_k}{\sum_l c_l} \quad (12)$$

We can now define probabilities similar to above, but within one of the cash flow groups *to* or *from* an agent for  $j \neq k$

$$p_k^{(j)} = \frac{c_k^j}{c^j} (1 - \delta_k^j) \quad p_{(k)}^j = \frac{c_k^j}{c_k} (1 - \delta_k^j). \quad (13)$$

This is the probability to find an arbitrary currency unit leaving/going to agent  $j$  in a cash flow to/from agent  $k$ . Evidently

$$\sum_k p_k^{(j)} = 1 \quad \sum_j p_{(k)}^j = 1 \quad (14)$$

By grouping the cash flow by the originating agent (upper index) we find

$$H_{\{c\}} = H' + \sum_j p^j H^j, \quad (15)$$

where

$$H' \stackrel{\text{def}}{=} - \sum_j p^j \log_2(p^j) \quad (16)$$

is the entropy of the aggregate agent cash out-flow probabilities  $p^j$  and

$$H^j \stackrel{\text{def}}{=} - \sum_k p_k^{(j)} \log_2(p_k^{(j)}) \quad (17)$$

is the entropy of each agents cash outflows, i.e., the entropy of  $p_k^{(j)}$  for fixed  $j$ . It is maximized if all cash flows from the agent are equal, and minimized (in which case it is zero) if there is only a single cash flow from agent  $j$ . Hence this is a measure of spending diversity for agent  $j$ .

Similarly this can be done for cash in-flows

$$H_{\{c\}} = H'' + \sum_k p_k H_k, \quad (18)$$

where  $H''$  is the entropy of the aggregate agent cash in-flow probabilities  $p_k$  and  $H_k$  is the entropy of the  $p_{(k)}^j$  for fixed  $k$ . Here similarly to above

$$H'' \stackrel{\text{def}}{=} - \sum_k p_k \log_2(p_k) \quad (19)$$

$$H_k \stackrel{\text{def}}{=} - \sum_j p_{(k)}^j \log_2(p_{(k)}^j). \quad (20)$$

Taking the sum and difference of the two expressions for  $H_{\{c\}}$  we find

$$H_{\{c\}} = \frac{H' + H''}{2} + \frac{1}{2} \sum_j (p_j H_j + p^j H^j) \quad (21)$$

$$0 = \frac{H'' - H'}{2} + \sum_j (p_j H_j - p^j H^j) \quad (22)$$

The last equation is particularly interesting as it relates the difference of overall income and spending inequality to individual agents cash flow diversification.

### 3 The Steady State Economy

Cash flows to and from agents over the period are related by

$$[c^1 \dots c^N] \begin{bmatrix} p_1^{(1)} & \dots & p_N^{(1)} \\ \vdots & & \vdots \\ p_1^{(N)} & \dots & p_N^{(N)} \end{bmatrix} = [c_1 \dots c_N] \quad (23)$$

For a stationary wealth distribution we need  $c_j = c^j$  and hence

$$p_j = p^j. \quad (24)$$

Assuming stationarity and dividing above equation by the sum of all cash flows gives

$$[p^1 \dots p^N] \begin{bmatrix} p_1^{(1)} & \dots & p_N^{(1)} \\ \vdots & & \vdots \\ p_1^{(N)} & \dots & p_N^{(N)} \end{bmatrix} = [p^1 \dots p^N] \quad (25)$$

In the case of stationarity we have  $\bar{H} \stackrel{\text{def}}{=} H' = H''$  and expression 21 becomes

$$H = \bar{H} + \frac{1}{2} \sum_j p^j (H_j + H^j), \quad (26)$$

which shows how the overall entropy splits into an inter-agent component  $\bar{H}$  and an internal component

for each agent  $\frac{H_j + H^j}{2}$ . We suggest that  $\bar{H}$  is a measure of income inequality as it measures the diversity of incomes. However,  $\bar{H}$  includes contributions from all agents, including non-human agents like corporations and government. Using the same identity as above one can express  $\bar{H}$  in terms of agent groups, for example

$$\bar{H} = H^* + p_G^* H_G^* + p_C^* H_C^* + p_P^* H_P^*, \quad (27)$$

where  $H_P^*$  measures the income distribution between human individuals and can be recognized as related to the Theil index of income inequality [2].  $H_G^*$  and  $H_C^*$  are similar measures for government and corporates.  $H^*$  then measures the inequality between the sectors in the economy. It is well known that these properties of the information entropy make it a suitable inequality measure [3].

Similarly (22) simplifies to

$$0 = \sum_j p^j (H_j - H^j), \quad (28)$$

which says that the probability weighted sum of the difference of agent inflow and outflow entropy vanishes. This equation is interesting as it shows how the entropy differentials between agent in and out flows have to balance in the economy.

For the agent in the economy labelled  $j$ , the entropy  $H_j$  represents the uncertainty of which source a dollar of  $j$ 's came from. Similarly, the entropy  $H^j$  is the uncertainty of where a dollar of  $j$ 's payments will go. The difference  $(H_j - H^j)$  is, therefore, the difference of uncertainty between the inflow and outflow probabilities. If the difference is positive, then an agent will be more uncertain about where an inflow dollar came from than where an outflow dollar is going. If the difference is negative then the opposite will be true.

Since the probabilities  $p^j$  in (28) are all positive, it is not possible for all agents to have a positive entropy difference or all to have a negative entropy difference. Some agents must be more certain about where their next dollar will come from relative to where it is going and other agents must be more certain about where their next dollar is going than where it came from. Typically, the former could be an individual (often

the whole income comes in the form of a single salary payment) while the latter could be a store or service provider (many customers pay for goods while the payments will go to few suppliers).

## 4 Examples

### 4.1 The two-agent economy

This is the simplest possible, but somewhat restricted case. Let us define the matrix of cash flows

$$C \stackrel{\text{def}}{=} \begin{bmatrix} c_1^1 & c_2^1 \\ c_1^2 & c_2^2 \end{bmatrix} \quad (29)$$

Note that there is the freedom to rescale the cash flows by a constant factor. For stationarity we require

$$c_2^1 = c_1^2. \quad (30)$$

With

$$\vec{u} \stackrel{\text{def}}{=} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (31)$$

we have

$$\begin{bmatrix} c_1^1 \\ c_2^1 \end{bmatrix} = C \vec{u} \quad (32)$$

$$\begin{bmatrix} p^1 \\ p^2 \end{bmatrix} = \frac{C \vec{u}}{\vec{u}' C \vec{u}} \quad (33)$$

$$\begin{bmatrix} p_1^{(i)} & p_2^{(i)} \end{bmatrix} = \begin{bmatrix} c_1^i & c_2^i \end{bmatrix} / (c_1^i + c_2^i) \quad (34)$$

From these all the entropies of interest ( $H'$ ,  $H_j$  and  $H^j$ ) can be calculated.

Let us scale the cashflows such that  $c_2^1 = c_1^2 = 1$ , i.e., the savings are measured relative to the inter-agent cashflows. The cashflow matrix then takes the form

$$C \stackrel{\text{def}}{=} \begin{bmatrix} a & 1 \\ 1 & b \end{bmatrix}, \quad (35)$$

where  $a$  and  $b$  are non-negative numbers. We then have

$$p_{\{c\}} = \frac{2}{2 + a + b} \quad p_{\{s\}} = 1 - p_{\{c\}} \quad (36)$$

$$H_{\{s\}} = K(a, b) \quad H_{\{c\}} = 1 \quad (37)$$

Figures 1 to 4 illustrate probabilities and entropies for this case.

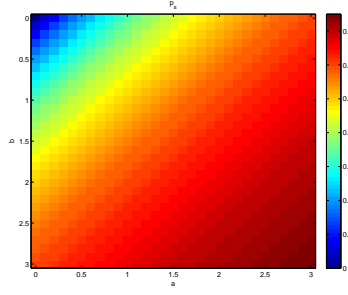


Figure 1: Heat-map of  $p_{\{s\}}$  for the 2-agent case.

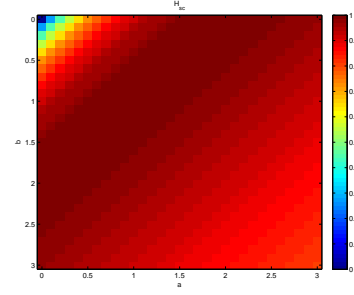


Figure 3: Heat-map of  $H_{\{sc\}}$  for the 2-agent case.

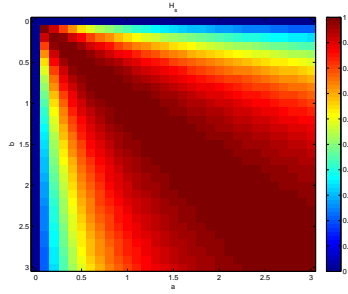


Figure 2: Heat-map of  $H_{\{s\}}$  for the 2-agent case.

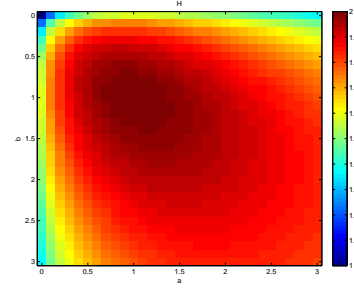


Figure 4: Heat-map of the overall cash-flow entropy  $H$  for the 2-agent case. The previous plots are a decomposition of this entropy. We have  $H = H_{\{sc\}} + pH_{\{s\}} + p_{\{c\}}H_{\{c\}}$ .

## 4.2 The three-agent economy

Let us assume stationarity and no savings. The cash-flow matrix is then

$$\begin{bmatrix} 0 & c_2^1 & c_3^1 \\ c_1^2 & 0 & c_3^2 \\ c_1^3 & c_2^3 & 0 \end{bmatrix} \quad (38)$$

There are two constraints resulting from the stationarity requirement

$$c^1 = c_1 \quad c^2 = c_2 \quad (39)$$

(this implies  $c^3 = c_3$ ) and one freedom to “choose the cash denomination” (allowing us to re-scale all values in the cash-flow matrix).

Let us choose the normalization such that the average agent income/spending is 1, i.e.,

$$c^1 + c^2 + c^3 = 3. \quad (40)$$

Furthermore, let us choose the following three parameters for the cash flow matrix

$$\frac{c_2^1}{c^1} = a \quad \frac{c_3^2}{c^2} = b \quad \frac{c_1^3}{c^3} = k. \quad (41)$$

Then the definition of  $c^j$  implies

$$\frac{c_3^1}{c^1} = 1 - a \quad \frac{c_1^2}{c^2} = 1 - b \quad \frac{c_2^3}{c^3} = 1 - k. \quad (42)$$

Writing down the definitions for  $c^1$ ,  $c^2$  and the mean constraint as a linear system in the  $c^j$  gives

$$\begin{bmatrix} c^1 \\ c^2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 & 1-b & k \\ a & 0 & 1-k \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} c^1 \\ c^2 \\ c^3 \end{bmatrix} \quad (43)$$

and hence

$$\begin{bmatrix} c^1 \\ c^2 \\ c^3 \end{bmatrix} = B^{-1} \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \quad (44)$$

where

$$B = \begin{bmatrix} -1 & 1-b & k \\ a & -1 & 1-k \\ 1 & 1 & 1 \end{bmatrix} \quad (45)$$

The overall entropy (which is equal to  $H_{\{c\}}$  because we excluded savings here) for different values of  $k$  is presented in the heat plots 5 and 6.

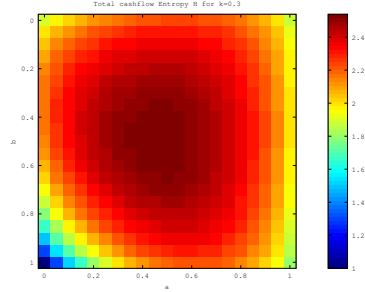


Figure 5: Heat-map of the overall cash-flow entropy  $H$  for the 3-agent case with  $k = 0.3$ .

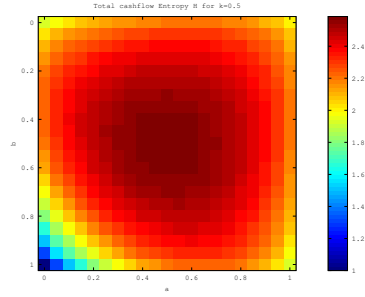


Figure 6: Heat-map of the overall cash-flow entropy  $H$  for the 3-agent case with  $k = 0.5$ .

To illustrate the identity (28) let us assume a specific case, by fixing the parameters for the cash flow matrix to

$$a = 0.1 \quad b = 0.3 \quad k = 0.7. \quad (46)$$

This corresponds to the cash flow matrix

$$\begin{bmatrix} 0 & 0.1235 & 1.1118 \\ 0.3516 & 0 & 0.1507 \\ 0.8837 & 0.3787 & 0 \end{bmatrix} \quad (47)$$

The inflow entropies (entropies in each column) are then

$$H_1 = 0.8617; \quad H_2 = 0.8048; \quad H_3 = 0.5275 \quad (48)$$

while the outflow entropies (entropies in each row) are then

$$H^1 = 0.4690; \quad H^2 = 0.8813; \quad H^3 = 0.8813. \quad (49)$$

Let us define the entropy differentials (inflow minus outflow entropy)  $D_j \stackrel{\text{def}}{=} H_j - H^j$

$$D_1 = 0.3927; \quad D_2 = -0.0765; \quad D_3 = -0.3538 \quad (50)$$

We see that agent 1 is an entropy reducer, while agent 3 is an entropy increaser. Agent 2 is almost entropy neutral. From the cash flow matrix we find the  $p_j$  as

$$p_1 = 0.4118; \quad p_2 = 0.1674; \quad p_3 = 0.4208. \quad (51)$$

Now it is easy to verify that the identity (28) is satisfied. This means that the weighted entropy differentials for any two agents imply the weighted entropy differential for the remaining agent.

## 5 Conclusion

We presented a formalism linking cash flows to an entropy measure. The measure naturally adopts to sub-groupings of the cash flows and as such produces a number of inequality measures, one of them being the Theil Index.

While the Theil Index is well known, we believe the over all framework we presented is new. It produces

inequality measures between all possible groupings of agents and internal inequality measures for each group.

Another interesting result is a kind of overall entropy balance expressed in equation (22) and (28). This essentially says that there must be a balance between entropy increasers (receiving few large cash flow, making many small payments) and decreasers (receiving many small payments and making a few large payments). Whether this can be linked to an “economic energy” is an interesting question which we have not yet addressed. However, one might speculate that entropy increasers (consumers) tend to pay the profit margins received by entropy decreasers (companies).

The formalism presented here does not allow for credit, which is obviously an important factor in the real world. We believe that credit can be incorporated in the formalism and this will be the subject of further research.

Another possible extension is an inclusion of time so that entropy cannot just be measured over one time step, but also over a number of time steps. One way of doing this is to consider discrete times and to treat the same agent at different times as ‘different’. This should lead to entropies for payment profiles over time, for example for loans repayed in installments. We hope to be able to explore this more in a separate paper.

## A Subdivision Identity Example

Let us assume that we have 2 possible outcomes labeled by 1 and 2 with probabilities  $p_1$  and  $p_2 = 1 - p_1$ . The information entropy is then

$$H = -p_1 \log(p_1) - p_2 \log(p_2). \quad (52)$$

Now let outcome 2 be the combination of two possible outcomes 21 and 22 with probabilities  $p_{21}$  and  $p_{22}$  such that  $p_2 = p_{21} + p_{22}$ . We then find for the over-

all entropy

$$\begin{aligned}
H &= -p_1 \log(p_1) - p_{21} \log(p_{21}) - p_{22} \log(p_{22}) \\
&= -p_1 \log(p_1) \\
&\quad - p_{21} \log\left(p_2 \frac{p_{21}}{p_2}\right) - p_{22} \log\left(p_2 \frac{p_{22}}{p_2}\right) \\
&= -p_1 \log(p_1) - p_2 \log(p_2) \\
&\quad + p_2 \left[ -\frac{p_{21}}{p_2} \log\left(\frac{p_{21}}{p_2}\right) - \frac{p_{22}}{p_2} \log\left(\frac{p_{22}}{p_2}\right) \right]
\end{aligned}$$

The first two terms are the information entropy for outcome 1 and the aggregate outcome 2. The second term then adds the internal information entropy resulting from the sub-division of outcome 2, which is the weighted information entropy of the probabilities conditional on the knowledge that outcome 2 is true.

Above derivation can be generalized in an obvious way to arrive at (5).

## References

- [1] Jaynes E T, Bretthorst L G (2003), *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge
- [2] Theil H (1967), *Economics and Information Theory*, North Holland Publishing Company, Amsterdam
- [3] Shorrocks A F (1980), ‘The Class of Additive Decomposable Inequality Measures’, *Econometrica* **Vol 48 No 3**, 613-25